

We claim:

1. A method of providing improved quality of service over a series of related messages exchanged between computers in a networking environment, comprising steps of:
 - 3 determining one or more transactional quality of service ("TQoS") values to be applied to the related messages;
 - 5 using the determined TQoS values to transmit at least one of the related messages for delivery to a particular one of the computers; and
 - 7 annotating selected ones of the related messages with information reflecting the determined TQoS values.

- 1 2. The method according to Claim 1, wherein one of the TQoS values is a transmission priority value to be used when transmitting the annotated messages.
- 1 2 3. The method according to Claim 1, wherein one of the TQoS values is available bandwidth information pertaining to a network connection to the particular computer.
- 1 2 3. The method according to Claim 1, further comprising the step of storing the determined TQoS values for use when transmitting subsequent ones of the related messages to the particular computer.
- 1 2 5. The method according to Claim 1, wherein the particular computer is a client computer and the using step transmits one of the annotated messages to the client computer, and further

3 comprising steps of:

4 receiving the transmitted annotated message at the client computer; and

5 automatically returning the TQoS values to a server computer in each subsequent one of
6 the related messages.

1 6. The method according to Claim 5, wherein the transmitted annotated message includes an
2 object reference that is annotated to carry the TQoS values, and wherein the automatically
3 returning step is enabled by the annotation of the object reference.

1 7. The method according to Claim 1, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer.

1 8. The method according to Claim 1, wherein at least one of the annotated messages is a
2 request from the particular computer for a Web page.

1 9. The method according to Claim 1, wherein at least one of the annotated messages is a
2 request from the particular computer for a Web object.

1 10. The method according to Claim 5, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer and wherein at least one of the
3 subsequent ones of the related messages is a request for information referenced by the Web page.

1 11. The method according to Claim 5, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer and wherein at least one of the
3 subsequent ones of the related messages is a request for information selected from the Web page
4 by a user of the particular computer.

1 12. The method according to Claim 1, wherein the using step further comprises using the
2 determined TQoS values to set markings in a network layer header of the transmitted annotated
3 messages.

1 13. The method according to Claim 3, further comprising the step of enforcing bandwidth
2 allocation using the available bandwidth information as the at least one transmitted message is
3 transmitted through the networking environment.

1 14. The method according to Claim 2, further comprising the step of using the transmission
2 priority value to prioritize the transmission of the at least one transmitted message through the
3 networking environment.

1 15. The method according to Claim 4, wherein the storing step stores the determined TQoS
2 values in a server computer.

1 16. The method according to Claim 2 or Claim 3, wherein the annotating step further
2 comprises storing the information reflecting the determined TQoS values as part of a routing

3 token in the annotated messages.

1 17. The method according to Claim 16, wherein the routing token is used to modify a Uniform
2 Resource Locator from a header of selected ones of the related messages.

1 18. The method according to Claim 17, wherein the routing token further comprises
2 information enabling identification of the particular computer and another computer which
3 performs the transmitting step, as well as identification of a storage area used to store the
4 determined TQoS values for the related messages.

1 19. A system for providing improved quality of service over a series of related messages
2 exchanged between computers in a networking environment, comprising:

3 means for determining one or more transactional quality of service ("TQoS") values to be
4 applied to the related messages;

5 means for using the determined TQoS values to transmit at least one of the related
6 messages for delivery to a particular one of the computers; and

7 means for annotating selected ones of the related messages with information reflecting the
8 determined TQoS values.

1 20. The system according to Claim 19, wherein the TQoS values comprise one or more of (1)
2 a transmission priority value to be used when transmitting the annotated messages and (2)
3 available bandwidth information pertaining to a network connection to the particular computer.

1 21. The system according to Claim 19, further comprising means for storing the determined
2 TQoS values for use when transmitting subsequent ones of the related messages to the particular
3 computer.

1 22. The system according to Claim 19, wherein the particular computer is a client computer
2 and wherein the means for using the determined TQoS values transmits one of the annotated
3 messages to the client computer, and further comprising:

4 means for receiving the transmitted annotated message at the client computer; and
5 means for automatically returning the TQoS values to a server computer in each
6 subsequent one of the related messages.

1 23. The system according to Claim 22, wherein the transmitted annotated message includes an
2 object reference that is annotated to carry the TQoS values, and wherein the means for
3 automatically returning is enabled by the annotation of the object reference.

1 24. The system according to Claim 19, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer, a request from the particular
3 computer for a Web page, or a request from the particular computer for a Web object.

1 25. The system according to Claim 22, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer and wherein at least one of the

3 subsequent ones of the related messages is a request for information referenced by the Web page.

1 26. The system according to Claim 22, wherein at least one of the annotated messages is a
2 response that serves a Web page to the particular computer and wherein at least one of the
3 subsequent ones of the related messages is a request for information selected from the served Web
4 page by a user of the particular computer.

1 27. The system according to Claim 19, wherein the means for using the determined TQoS
2 values further comprises using the determined TQoS values to set markings in a network layer
3 header of the transmitted annotated messages.

1 28. The system according to Claim 20, further comprising means for enforcing bandwidth
2 allocation using the available bandwidth information as the at least one transmitted message is
3 transmitted through the networking environment.

1 29. The system according to Claim 20, further comprising means for using the transmission
2 priority value to prioritize the transmission of the at least one transmitted message through the
3 networking environment.

1 30. The system according to Claim 21, wherein the means for storing stores the determined
2 TQoS values in a server computer.

1 31. The system according to Claim 20, wherein the means for annotating further comprises
2 means for storing the information reflecting the determined TQoS values as part of a routing
3 token in the annotated messages.

1 32. The system according to Claim 31, wherein the routing token is used to modify a Uniform
2 Resource Locator from a header of selected ones of the related messages.

1 33. The system according to Claim 32, wherein the routing token further comprises
2 information enabling identification of the particular computer and another computer which
3 performs the means for transmitting, as well as identification of a storage area used to store the
4 determined TQoS values for the related messages.

1 34. The system according to Claim 22, wherein:

2 the TQoS values comprise at least (1) a transmission priority value to be used when
3 transmitting the annotated messages and (2) available bandwidth information pertaining to a
4 network connection to the particular computer; and

5 at least one of the annotated messages is a response that serves a Web object to the
6 particular computer from a network cache; and

7 wherein the means for using the determined TQoS values further comprises using the
8 determined TQoS values, by the network cache, to prioritize transmission of the response that
9 serves the Web object and to enforce bandwidth allocation using the available bandwidth
10 information as the response is transmitted.

1 35. A system for providing improved quality of service for transmission of related request and
2 response messages exchanged between computers in a networking environment, comprising:

3 means for determining one or more quality of service ("QoS") values to be applied to
4 transmission of the related messages; and

5 means for communicating the QoS values to be applied to the transmission by storing the
6 determined QoS values in headers of selected ones of the request and response messages.

1 36. The system according to Claim 35, wherein the determined QoS values are stored as
2 cookies in the headers.

1 37. A computer program product for providing improved quality of service over a series of
2 related messages exchanged between computers in a networking environment, the computer
3 program product embodied on one or more computer-readable media and comprising:

4 computer-readable program code means for determining one or more transactional quality
5 of service ("TQoS") values to be applied to the related messages;

6 computer-readable program code means for using the determined TQoS values to transmit
7 at least one of the related messages for delivery to a particular one of the computers; and

8 computer-readable program code means for annotating selected ones of the related
9 messages with information reflecting the determined TQoS values.

1 38. The computer program product according to Claim 37, wherein the TQoS values

2 comprise one or more of (1) a transmission priority value to be used when transmitting the
3 annotated messages and (2) available bandwidth information pertaining to a network connection
4 to the particular computer.

1 39. The computer program product according to Claim 37, further comprising computer-
2 readable program code means for storing the determined TQoS values for use when transmitting
3 subsequent ones of the related messages to the particular computer.

1 40. The computer program product according to Claim 37, wherein the particular computer is
2 a client computer and wherein the computer-readable program code means for using the
3 determined TQoS values transmits one of the annotated messages to the client computer, and
4 further comprising:

5 computer-readable program code means for receiving the transmitted annotated message
6 at the client computer; and

7 computer-readable program code means for automatically returning the TQoS values to a
8 server computer in each subsequent one of the related messages.

1 41. The computer program product according to Claim 40, wherein the transmitted annotated
2 message includes an object reference that is annotated to carry the TQoS values, and wherein the
3 computer-readable program code means for automatically returning is enabled by the annotation
4 of the object reference.

1 42. The computer program product according to Claim 37, wherein at least one of the
2 annotated messages is a response that serves a Web page to the particular computer, a request
3 from the particular computer for a Web page, or a request from the particular computer for a Web
4 object.

1 43. The computer program product according to Claim 40, wherein at least one of the
2 annotated messages is a response that serves a Web page to the particular computer and wherein
3 at least one of the subsequent ones of the related messages is a request for information referenced
4 by the Web page.

1 44. The computer program product according to Claim 40, wherein at least one of the
2 annotated messages is a response that serves a Web page to the particular computer and wherein
3 at least one of the subsequent ones of the related messages is a request for information selected
4 from the served Web page by a user of the particular computer.

1 45. The computer program product according to Claim 37, wherein the computer-readable
2 program code means for using the determined TQoS values further comprises using the
3 determined TQoS values to set markings in a network layer header of the transmitted annotated
4 messages.

1 46. The computer program product according to Claim 38, further comprising computer-
2 readable program code means for enforcing bandwidth allocation using the available bandwidth

3 information as the at least one transmitted message is transmitted through the networking
4 environment.

1 47. The computer program product according to Claim 38, further comprising computer-
2 readable program code means for using the transmission priority value to prioritize the
3 transmission of the at least one transmitted message through the networking environment.

1 48. The computer program product according to Claim 39, wherein the computer-readable
2 program code means for storing stores the determined TQoS values in a server computer.

1 49. The computer program product according to Claim 38, wherein the computer-readable
2 program code means for annotating further comprises computer-readable program code means for
3 storing the information reflecting the determined TQoS values as part of a routing token in the
4 annotated messages.

1 50. The computer program product according to Claim 49, wherein the routing token is used
2 to modify a Uniform Resource Locator from a header of selected ones of the related messages.

1 51. The computer program product according to Claim 48, wherein the routing token further
2 comprises information enabling identification of the particular computer and another computer
3 which performs the computer-readable program code means for transmitting, as well as
4 identification of a storage area used to store the determined TQoS values for the related

messages.